

# Estadística multivariada, 1 sem. 2019

Juan Carlos Castillo & Alejandro Plaza

Sesión 1: Introducción

Preambulo: ¿Cómo explicamos fenómenos sociales?

# Pedro, Juan ...

**Pedro** estudió en un colegio particular pagado del barrio alto de Santiago, y fue un buen alumno. En el Simce de 8° básico sacó más de 316 puntos, lo que lo situó en el 10% más alto entre todos los niños chilenos.

**Juan** también superó los 316 puntos. En su caso, el mérito fue doble, porque estudió en un colegio público: el capital cultural y el nivel socioeconómico de su familia eran muy inferiores a los de Pedro. Citando una famosa metáfora, Pedro corrió con patines mientras Juan iba descalzo, y así y todo llegaron empatados a la meta.

**¿Cómo premiará la sociedad el mérito de Pedro y el de Juan? ¿Reconocerá más a quien, contra todo pronóstico, logró destacarse pese a un entorno desfavorable?**

Gracias a una investigación hasta ahora inédita, hoy podemos responder esa pregunta (Meneses, Blanco & Paredes): quienes tuvieron más de 316 puntos en el Simce 2004, y estudiaban en uno de nueve colegios de élite de Santiago (nuestros “Pedros”), en 2017 ganaron un sueldo bruto promedio de \$1.471.115. Quienes tuvieron el mismo resultado en el Simce, pero estaban en un liceo público (nuestros “Juanes”), recibieron \$716.665.

# Pedro, Juan ... & Diego

Hablemos ahora de **Diego**. Él, a diferencia de Pedro y Juan, fue un mal alumno. Pese a asistir a un colegio de elite, quedó por debajo del promedio nacional, con entre 217 y 251 puntos en el Simce. Pues bien, a pesar de sus malos resultados, en 2017 Diego ganaba \$1.071.756.

Juan partió la carrera con todas las desventajas: menor nivel socioeconómico y capital cultural, y una inversión por alumno entre 5 y 6 veces inferior a la que recibió Diego. Así y todo, en una hazaña improbable, le sacó más de 100 puntos de ventaja en el Simce. Pero, lejos de ser premiado por su talento sobresaliente, ahora Juan tiene un trabajo peor pagado que Diego.

(adaptado de columna de Daniel Matamala en La Tercera, ver original [aquí](#) )

# Explicación ... y un mundo de datos

- Aumento de bases de datos en diferentes temas, con mediciones comparables a nivel internacional, y con crecientes grados de complejidad. **Ejemplos.**
- Del mismo modo existe una creciente profesionalización en repositorios de datos, y curatoria enfocada a compartir, preservar, citar, explorar y analizar datos de otras investigaciones. **Proyecto Dataverse.**

# Este curso

$$y \leftarrow bx + e$$

(programa & sesiones)

# El programa

# Definiciones y acuerdos básicos

- Estadística vs. investigación aplicada
- Orientación práctica: aprender haciendo
- Protagonismo en el proceso de aprendizaje
- Reproducibilidad, colaboración y ciencia abierta



## Parte 2: La explicación en sociología empírica

# Teorías y explicación

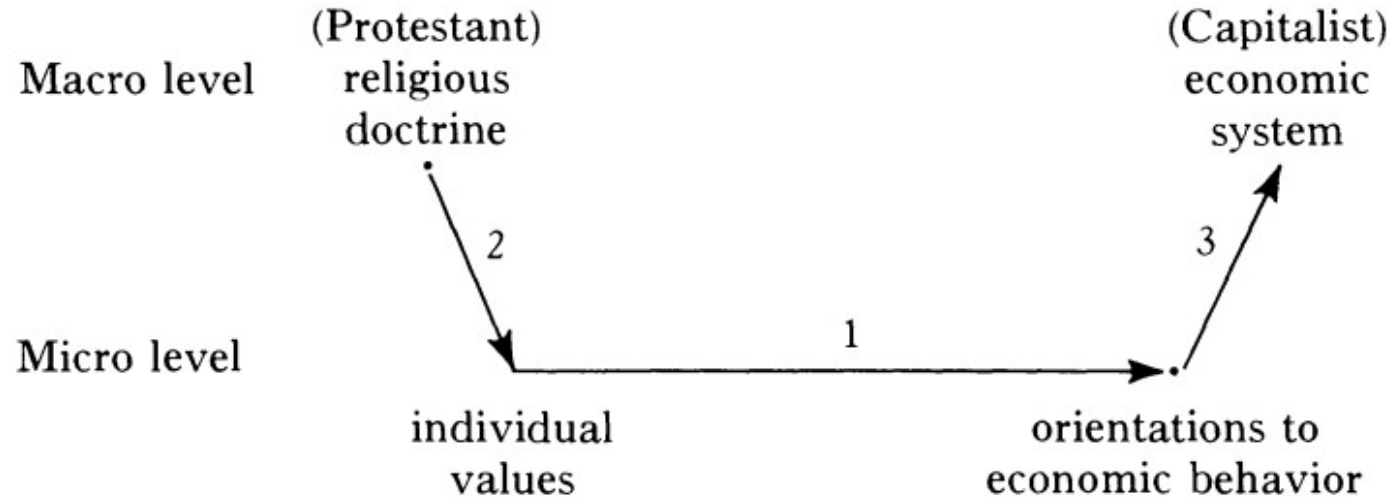


FIG. 2.—Macro-micro-macro relations: methodological individualism

# Explanandum y explanans

- **Explanandum**: el fenómeno que pretendemos explicar
  - definición precisa
  - relevancia
  - variabilidad
- **Explanans**: lo que genera la aparición del fenómeno
  - relación lógica (silogismo y "falacia de la afirmación del consecuente")
  - eficacia explicativa (elegancia, potencial heurístico, parsimonia)
  - claridad

# Algunos antecedentes de la explicación en sociología empírica

- Uso de métodos cuantitativos en sociología: rezago frente a disciplinas como la economía (econometría) y psicología (psicometría)
- Vaivenes y discontinuidades entre Europa y Estados Unidos
- Gustav von Schmoller y el Methodenstreit (deductivo vs inductivo): promueve extensivo uso de datos históricos, que confusamente se asocian a "causas".
- Fundación del laboratorio de estadística de universidad de Columbia: Giddings, "Statistics & Sociology" & Chapin, "Fieldwork & Social Research" (1920).

# Algunos antecedentes de la explicación en sociología empírica

- Uso de tablas y comienzos de regresión (Lundberg, 1929), basadas en nociones del positivismo lógico: búsqueda de definiciones operacionales genuinas y clasificaciones objetivas. Foco en correlación y abandono de causalidad.
- Stouffer et. al ("The American Soldier", 1949): estudio de actitudes basado en encuestas, y aporte de Lazarsfeld (1950) sobre control estadístico de relaciones espúreas mediante regresión múltiple.
- Etapa actual: relaciones múltiples entre variables (ecuaciones estructurales), foco en mecanismos y experimentos sociológicos. Vuelta a la relación entre explicación y causalidad.

# Modalidades de explicación en ciencias sociales (Linares,2018).

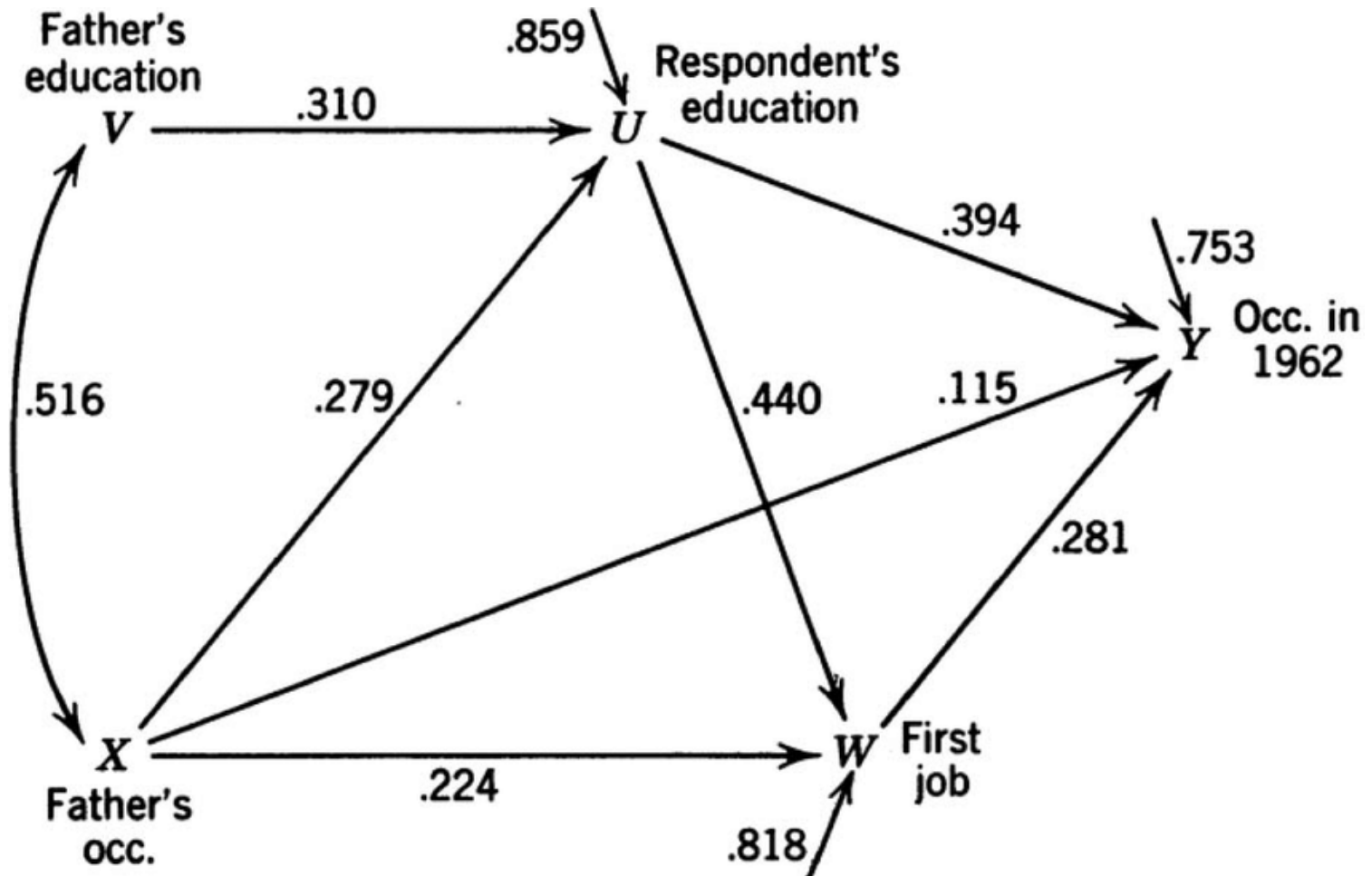
- Por leyes de cobertura
- Explicación Funcional
- Explicación Estadística
- Explicación "como si"
- *Explicación por mecanismos*

# Explicación y modelos estadísticos

¿Es posible lograr explicaciones desde el análisis de datos?

- existen mejores y peores formas de aproximarse a la explicación
- desde la regresión, una de las claves es el *control estadístico*, que permite lidiar con explicaciones alternativas

## ...Volviendo a Pedro, Juan & Diego





# Preguntas

- Responder anónimamente las siguientes preguntas en una hoja
  - Solo anotar el número de la pregunta y la respuesta
  - Si no se sabe la respuesta, simplemente escribir "No sé"
- 

1. ¿Entre qué rango varía una correlación?

2. ¿Para qué se usa el test de diferencia de  $\chi^2$ ?

3. ¿Cual es la relación entre varianza y desviación estándar?

# Estadística multivariada - Sesión 1